Idempotent Binary–>Decimal–>Binary Conversion

Suppose binary floating-point carries p sig. bits, and floating-point decimal strings are put out with P sig. dec. How big a value P suffices to ensure that correctly rounded conversion from binary to decimal and then from decimal back to binary recreates the original binary number?

Consider a real number x in a *Binade* $2^B \le x \le 2^{B+1}$ and in a *Decade* $10^D \le x \le 10^{D+1}$ where B and D are suitable integers; this implies that $2^B \le 10^{D+1}$ and $10^D \le 2^{B+1}$. The gap between adjacent binary floating-point numbers near x is 2^{B+1-p}; the gap between adjacent floating-point decimal numbers near x is 10^{D+1-P}. Conversion from binary to decimal will incur a rounding error no bigger than $5\cdot10^{D-P}$, and then conversion back to binary will incur an additional rounding error no bigger than 2^{B-p}. So long as these two rounding errors add up to less than the gap between adjacent binary numbers, the original number must be recreated; this means that when P is so big that $5\cdot10^{D-P} + 2^{B-p} < 2^{B+1-p}$ then P is big enough. This last inequality requires $P > D+1 - (B+1-p)\cdot\log_{10}2$. It must be satisfied when $P > 1 + p\cdot\log_{10}2$ because, as we saw above, $D \le (B+1)\cdot\log_{10}2$. Therefore P is sufficiently big when

$$P \ge \overline{P} := \text{ceil}(\ 1 + p\cdot\log_{10}2\ ) = \text{ceil}(\ 1 + p\cdot0.30103\dots) .$$

For instance, 8-byte wide double-precision floating-point numbers have precision $p = 53$, for which apparently a sufficiently big $P = 17$, barely bigger than $16.9\ldots = 1 + p\cdot0.30103\ldots$. No smaller P suffices, as can be verified by converting binary numbers barely less than 1024 .

The converse problem, so to speak, is to determine how *small* a value P suffices to ensure that correctly rounded conversion from decimal to binary and then from binary back to decimal recreates the original decimal number. Reasoning like that above implies that a sufficiently small

$$P \le \underline{P} := \text{floor}(\ (p-1)\cdot\log_{10}2\ ) = \text{floor}(\ (p-1)\cdot0.3010299\dots\ ) .$$

For instance, when $p = 53$ then $P = 15$ is small enough but 16 is not, as examples barely less than 0.001 reveal. Thus, for $p = 53$ sig. bits, the idempotent (reproducing) conversions are

Binary–>Decimal–>Binary when $P \ge \overline{P} := 17$ sig. dec.,
Decimal–>Binary–>Decimal when $P \le \underline{P} := 15$ sig. dec.

The difference between $\underline{P} = 15$ and $\overline{P} = 17$ is unusually small. For different binary precisions the differences are bigger:

For single-precision binary	$p = 24$,	the decimal precisions are	$\underline{P} = 6$ and $\overline{P} = 9$.
… double-extended	$p = 64$		$\underline{P} = 18$ and $\overline{P} = 21$.
… quadruple-precision	$p = 113$		$\underline{P} = 33$ and $\overline{P} = 36$.

The difference between $\underline{P}$ and $\overline{P}$ can be narrowed by sufficiently restricting the range of numbers x being converted, but that is a story for another day.